Journal of Agrometeorology

(A publication of Association of Agrometeorologists)

ISSN : 0972-1665 (print), 2583-2980 (online)
Vol. No. 27 (4) : 511-513 (December - 2025)
https://doi.org/10.54386/jam.v27i4.3151
https://journal.agrimetassociation.org/index.php/jam

Short communication

Comparative analysis of weather-driven models for sorghum yield prediction in Bundelkhand

RITU SINGH[1], ANNU[2], UMESH CHANDRA[1], HIMANI MAHESHWARI[4], ARJUN PRASAD VERMA[3] and GAURAV SHUKLA[1]

[1]Department of Statistics & Computer Science, Banda University of Agriculture & Technology, Banda, Uttar Pradesh, India
[2]Department of Basic & Social Science, Banda University of Agriculture & Technology, Banda, Uttar Pradesh, India
[3]Department of Agricultural Extension, Banda University of Agriculture & Technology, Banda, Uttar Pradesh, India
[4]School of Computing, Graphic Era Hill University, Dehradun, Uttarakhand, India
Corresponding author's e-mail: gauravshuklastat@gmail.com

Bundelkhand region of Uttar Pradesh and Madhya Pradesh states faces various problems such as continuous drought, poor management of rainfall, soil loss, single crop focus, poverty and continuous migration of people out of the area. Agriculture in this area is based on rain fed crops such as pulses and oilseeds (Kalia *et al.,* 2021). With its ability to tolerate dry spells and prosper in harsh soil conditions, sorghum is most suited cereal crop for Bundelkhand's semi-arid land where rainfall is uncertain, and water access is scarce. Sorghum is a vital basic food for rural families. It is extremely sensitive to temperature and photoperiod; it is important to study the role of meteorological variables to improve sorghum production and its quality. India is the third- largest producer of sorghum in the world with 6.0 million metric tons (9% of global production) after United States (14% of global production) and Nigeria (10% of global production) (IPAD 2024-2025). Agrometeorological indicators are vital for adapting agricultural planning, which enhances crop management and promotes food security under threat due to climate change. In the past, researchers have employed important statistical methods for crop harvesting challenges due to weather condition and its prediction on pre harvest forecasting (Garde *et al.,* 2015; Maurya *et al.,* 2025). Machine learning models employ algorithms to assess data, spot trends, and develop connects within the dataset, as compared to statistical models that link weather data and agricultural output using pre-established mathematical equations. Machine learning models have been successfully used to predict the yields of a number of crops including cashew, pigeon pea, rice, sorghum, coconut, wheat and soybean (Satpathi *et al.,* 2023; Khan *et al.,* 2023; Setiya *et al.,* 2022).

The study focuses on two major sorghum (jowar) producing districts of Uttar Pradesh, Banda (25^{0}48’ N, 80^{0}33’ E) and Hamirpur (25^{0}95’ N,80^{0}15’E). Both districts are recognized for being characterized by fertile alluvial and black soils with a high capacity for holding moisture and their area is suitable for growing sorghum because of its favorable soil characteristics and agro-climatic factors (Suman *et al.,* 2024). Time series data on sorghum yield for about 31 years from 1991-92 to 2021-22, especially for the Banda and Hamirpur districts of Uttar Pradesh, was obtained from the Spider Statistics Journal, published by the Directorate of Economics and Statistics, Uttar Pradesh. The data on weather conditions such as maximum temperature (Tmax), minimum temperature (Tmin), rainfall (Rain), wind speed (WS), and relative humidity (RH) was gathered from NASA POWER web portal (https://power.larc.nasa.gov/data-access-viewer/). The weather variables data have been transformed into phenological stage wise averages to process further. In terms of phenology, the average values were computed using the daily weather data. These average values are subsequently employed in the computation of both weighted and unweighted weather indices as per procedure given by Setiya *et al.,* (2022). Using time series data on phenological weather indices and sorghum yield, three multivariate models viz. Artificial Neural Network (ANN), Principal Component Analysis (PCA-ANN) and Least Absolute Shrinkage and Selection Operator (LASSO) were developed for this study. The models were tested and trained on these datasets. Among the complete dataset spanning 31 years, 25 years of data were employed for training of the models, while the remaining 6 years data were utilized for testing of the models. In training stage, the description of the models with their error percentage are given in Table 1. The ANN model consists of an

***Article info - DOI**: https://doi.org/10.54386/jam.v27i4.3151*
Received:7 August 2025; Accepted:12 October 2025; Published online : 1 December 2025

Table 1: Description of the models used and percent error in Banda and Hamirpur districts

District	Model	Equation	Percent error (%)
Banda	ANN	No of hidden layers: 5	-20.7 - 8.1
	PCA-ANN	9; No of PC's: 5	-117.6 - 3
	LASSO	$Y = 8.438 + 1.02 \times Z11 + 0.148 \times Z251 + 0.666 \times Z351$	-53- 18.6
Hamirpur	ANN	No of hidden layers: 5	-120.7 - 16.5
	PCA-ANN	9; No of PC's: 5	-80.8 - 18.8
	LASSO	$Y = 7.486 + 0.513 \times Z11 + 0.269 \times Z31 + 0.18\times Z251 + 0.206 \times Z341$	-195.8 - 20.3

Note: Y = Predicted yield (q ha^{-1}); Z11 = weighted WS, Z31= weighted Tmin, Z251= weighted Tmin*Rain, Z351= weighted WS*Rain, Z341 = weighted WS*RH

Table 2: Quantitative measures obtained during training and testing stages for sorghum yield prediction models in two districts

Districts	Quantitative measures	Training			Testing		
		ANN	PCA-ANN	LASSO	ANN	PCA-ANN	LASSO
Banda	Adjusted R^2	0.86	0.93	0.70	0.90	0.99	0.40
	MAE (q ha^{-1})	0.65	0.35	1.00	1.35	0.18	8.5
	RMSE (q ha^{-1})	0.85	0.57	1.25	1.94	0.25	9.44
	nRMSE (%)	7.0	5.0	15.0	13.0	1.7	55.0
	EF	0.85	0.93	0.67	0.82	0.99	-3.13
Hamirpur	Adjusted R^2	0.87	0.94	0.51	0.94	0.85	0.83
	MAE (q ha^{-1})	0.58	0.33	1.22	0.38	0.38	1.46
	RMSE (q ha^{-1})	0.76	0.50	1.58	0.46	0.58	1.58
	nRMSE (%)	8.0	5.5	21.0	10.0	12.0	17.0
	EF	0.86	0.94	0.42	0.89	0.83	-0.21

input layer, five hidden layers and an output layer representing yield. In PCA-ANN model, PCA was first applied to the original dataset to extract five principal components (PCs) and these PCs were used as inputs to an ANN model with nine neurons and five hidden layers. LASSO is a regression technique used for variable selection and regularization in regression analysis. The model performance were assessed using adjusted R^2 (coefficient of determination), RMSE (root mean square error), mean absolute error (MAE), percentage error (PE), normalized root mean square error (nRMSE) and model efficiency (EF).

Artificial neural network (ANN)

Results of the analysis (Table 2) showed that the performance of Artificial Neural Network (ANN) was excellent for both Banda and Hamirpur districts with adjusted R^2 values of 0.86 and 0.87 along with RMSE 0.85 q ha^{-1} and 0.76 q ha^{-1} respectively during training, while during testing stage R^2 was 0.90 and 0.94 along with RMSE of 1.94 q ha^{-1} and 0.46 q ha^{-1} respectively. The value of nRMSE for Banda and Hamirpur district was less than 10%, indicating excellent model performance in training stage but in testing condition it was found 13% and 12% respectively. The modelling efficiency was 0.82 for Banda and 0.83 for Hamirpur district. The MAE value close to zero for both the districts indicate good model performance at training stage, though at the testing stage it shows the under estimation of the crop yield for both the districts.

PCA-ANN

The PCA-ANN model had the best performance during the training phase in Banda district (Table 2), with an adjusted R^2 of 0.93 along with lowest errors, including MAE (0.35 q ha^{-1}), RMSE (0.57 q ha^{-1}) and nRMSE (5%). The performance of PCA-ANN during testing was impressive with an adjusted R^2 of 0.99, with very low errors: MAE of 0.18 q ha^{-1}, RMSE of 0.25 q ha^{-1}, and nRMSE of just 1.7%. In Hamirpur district, PCA-ANN model continued to perform best. It had a high adjusted R^2 of 0.94, matching its performance during training, which shows it was consistent. The errors remained low, with MAE (0.33 q ha^{-1}), RMSE (0.50 q ha^{-1}) and nRMSE at 5.5%, proving it predicted yields very well. The efficiency factor (EF) in Banda was higher (0.99) than Hamirpur (0.83). Despite a good model performance at the training stage, the MAE value indicated underestimation of the crop yield at the testing stage for Hamirpur district.

Least absolute shrinkage and selection operator (LASSO)

The data on LASSO analysis of sorghum crop yield for training stage revealed that the value of adjusted R^2 was observed 0.70 for Banda district with RMSE value of 1.25 q ha^{-1}, whereas it was observed 0.51 for Hamirpur with RMSE value of 1.58 q ha^{-1}. The LASSO model didn't perform well at all in testing, showing an adjusted R^2 of 0.40, and high error values, MAE of 8.5 q ha^{-1}, RMSE of 9.44 q ha^{-1}, and nRMSE of 55% for Banda while for Hamirpur, it's performance also declined a lot during testing, with a low adjusted R^2 of 0.83 and higher errors MAE of 1.46 q ha^{-1}, RMSE of 1.58 q ha^{-1} and nRMSE of 17%. Its EF was also quite low at -3.13 and -0.21 for Banda and Hamirpur respectively, which confirms its weak predictive ability.

It is evident from results that PCA-ANN outperformed the other models in both districts. PCA-ANN had the highest adjusted R^2 (0.93) followed by ANN (0.86) and LASSO (0.70) during the testing for Banda district. The same pattern showed up in Hamirpur, where PCA-ANN again outperformed the others and was the most accurate on all the metrics. When, we moved to the testing phase, PCA-ANN kept its top spot for both the districts. In Banda, it had an impressive Adjusted R^2 of 0.99, lowest errors (MAE = 0.18 q ha^{-1}, and RMSE= 0.25 q ha^{-1}, nRMSE= 1.7%), showing that it is quite reliable. The regular ANN also did well, with an adjusted R^2 of 0.90 and EF of 0.82. LASSO, however, performed poorly during testing, especially in Banda, where it even had an adjusted R^2 of 0.40, and very high error values, this suggests it might have overfitted to the training data. Similar results were seen in Hamirpur, where PCA-ANN remained the most accurate model. The weak performance of LASSO during testing in both districts shows that linear models struggle to capture the complex, nonlinear relationships in weather and crop yield data. ANN is nonlinear by nature; it handled these relationships better. The hybrid PCA-ANN model did even better since it reduces dimensionality and filters out noise in the input data. This shows that ANN-based models perform better than traditional regression techniques (Aravind *et al.,* 2022 and Setiya *et al.,* 2022). The performance of these models was probably influenced by the regions' weather. Considering the results of PCA-ANN in the present study, it appears that hybrid approaches are more helpful in addressing complex agroclimatic factors (erratic rainfall patterns and periodic droughts) to enhance accuracy of yield predictions.

The present study examined the predictive ability of various models for sorghum yields in Banda and Hamirpur districts of Uttar Pradesh. Individual models such as ANN and LASSO, as well as a combination model dubbed PCA-ANN, were tested. On the whole, the results indicate that PCA-ANN is the best model that can be used to predict yield in the different agro-climate areas, with ANN ranking next, and LASSO ranking third and perhaps not appropriate in tracking the complexity of sorghum yields in these regions.

ACKNOWLEDGEMENT

The authors sincerely thank the anonymous reviewers for their careful reading, constructive comments and valuable suggestions, which greatly improved the quality and clarity of this manuscript.

Funding: No external funding is involved in this study.

Conflict of Interest: The authors declare no competing interests.

Data Availability: To be provided on request.

Author's Contribution: **R Singh:** Investigation, Formal analysis; **Annu:** Writing-review, Editing, Writing-original draft; **U. Chandra:** Conceptualization, Methodology, Writing; **H. Maheshwari**: Conceptualization, Methodology, Writing; **A.P. Verma**: Writing-review; **G. Shukla:** Supervision, Writing-review.

***Disclaimer*:** The contents, opinions, and views expressed in the research article published in the Journal of Agrometeorology are the views of the authors and do not necessarily reflect the views of the organizations they belong to.

***Publisher's Note*:** The periodical remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

REFERENCES

Aravind, K.S., Vashisth, A., Krishanan, P. and Das, B. (2022). Wheat yield prediction based on weather parameters using multiple linear, neural network and penalised regression models. *J. Agrometeorol.,* 24(1): 18-25. https://doi.org/10.54386/jam.v24i1.1002

Garde, Y.A., Dhekale, B.S. and Singh, S. (2015). Different approaches on pre harvest forecasting of wheat yield. *J. Appl. Natural Sci.,* 7(2): 839-843.

IPAD (2024–25). Sorghum yield and production position in the world**.** New Delhi: Indian Production and Area Database, Ministry of Agriculture & Farmers Welfare, Government of India.

Kalia, A., Shukla, G., Mishra, D., Mishra, B.P. and Patel, R.R. (2021). Comparative trend analysis of mustard in Bundelkhand region, Uttar Pradesh and India, *Indian J. Ext. Edu.,* 57(1): 15-19.

Khan, Y., Kumar, V., Setiya, P. and Satpathi, A. (2023). Comparison of phenological weather indices based statistical, machine learning and hybrid models for soybean yield forecasting in Uttarakhand. *J. Agrometeorol.,* 25(3): 425-431. https://doi.org/10.54386/jam.v25i3.2232

Maurya, A. K., Rai, R. K., Gautam, Y., Ojha, P.K., Shukla, G. and Pushpa (2025). Forecasting of Gram Production in Bundelkhand Region: A Time Series Analysis Using ARIMA Model. *J. Sci. Res. Rep.,* 31 (9): 288-298.

Satpathi, A., Setiya, P., Das, B., Nain, A.S., Jha, P.K., Singh, S. and Singh, S. (2023). Comparative analysis of statistical and machine learning techniques for rice yield forecasting for Chhattisgarh, India. *Sustainability,* 15(3): 2786.

Setiya, P., Satpathi, A., Nain, A.S. and Das, B. (2022). Comparison of weather-based wheat yield forecasting models for different districts of Uttarakhand using statistical and machine learning techniques. *J. Agrometeorol.,* 24(3): 255-261. https://doi.org/10.54386/jam.v24i3.1571

Suman, A. S., Mishra, A., Shukla, G., Sah, D., Chandra, U., Chaubey, A.K., Mishra, B.P., Pathak, J. and Panwar, G.S. (2024). Analyzing Alternatives for Managing Nitrogen in Puddled Transplanted Rice in a Semi-Arid Area of India. *Sustainability,* 16: 6096-6111.